Discrete Applied Mathematics 12 (1985) 1-11
North-Holland

MAXIMIZING A SUPERMODULAR PSEUDOBOOLEAN FUNCTION: A POLYNOMIAL ALGORITHM FOR SUPERMODULAR CUBIC FUNCTIONS

A. BILLIONNET
IIE/CNAM and Département de Mathématiques Appliquées du CNET, Paris, France

M. MINOUX
Département de Mathématiques Appliquées du CNET, Paris, France

Received 19 January 1984
Revised 20 November 1984

The problem of maximizing a pseudoboolean function (or equivalently a set function) which is supermodular, has many interesting applications e.g. in combinatorial optimization, Operations Research etc. Up to now, a number of special cases of pseudoboolean functions have been known, the maximization of which can be converted into the search for a maximum flow in an associated network. These were essentially the so-called *negative–positive* pseudoboolean functions (which, as will be noted here, turn out to be supermodular). First it is shown here how these results on *negative–positive* functions can be more easily derived by using the concept of *conflict graph*. The *conflict graph* approach is then generalized to extend the class of problems amenable to maximum network flow problems to the whole set of cubic supermodular pseudoboolean functions.

1. Introduction

Many problems in graph theory or combinatorial optimization can be formulated as the search for the maximum or the minimum of a *pseudoboolean* function, i.e. a real function of n binary variables $x_1, x_2, \dots, x_n$, which, without loss of generality, may be written as:

$$f(x_1, x_2, \dots, x_n) = \sum_{k=1}^{p} b_k \prod_{j\in N_k} x_j + K$$

where, for any k $(1 \le k \le p)$, b_k is a real number, N_k is the index subset of the variables in the kth monomial, and K is a constant.

We note that function f may be equivalently considered as a *set function* $\mathscr{P}(E) \to \mathbb{R}$ on $E = \{1, 2, \dots, n\}$ by considering any n-vector $x \in \{0,1\}^n$ as the characteristic vector of the subset $S_x = \{i \mid i \in E;\ x_i = 1\}$.

Optimizing an arbitrary pseudoboolean function (or equivalently an arbitrary set function) belongs to the class of the so-called NP-complete problems [10] for which no polynomial algorithms are known (and probably do not exist). There is, however, a number of interesting special cases for which polynomial algorithms

0166-218X/85/$3.30

could be found, in particular the so-called *negative–positive* pseudoboolean functions (where the coefficients of the linear terms have any sign, and higher degree terms have positive coefficients) the maximization of which can be converted into a maximum network flow computation in an associated graph [18].

A situation which generalizes the above special case is when function f to be maximized enjoys the so-called *supermodularity property*. In terms of set functions, a function $f: \mathscr{P}(E) \to \mathbb{R}$ is said to be supermodular if and only if, for any $A \subset E$, $B \subset A$, $i \notin A$:

$$f(A \cup \{i\}) - f(A) \geq f(B \cup \{i\}) - f(B).$$

Other equivalent definitions of supermodularity will be found in [8] for instance.

Our purpose here is to show that the class of problems amenable to maximum network flow computations can be extended to the whole set of cubic supermodular pseudoboolean functions (which is not limited to cubic negative–positive pseudoboolean functions). This result may be considered as a step towards the search for polynomial algorithms, more efficient in practice than Khachian's algorithm for maximizing an arbitrary supermodular pseudoboolean function (cf. Grötschel, Lovász & Schrijver [12]).

2. A special case of supermodular functions: negative–positive pseudoboolean functions

Let $f(x_1, \ldots, x_n)$ be a pseudoboolean function of n binary variables x_i $(i = 1, 2, \ldots, n)$. For any $i \in E$, $j \in E$, $i \neq j$ define the pseudoboolean functions φ_{ij} and ψ_{ij} by:

$$\begin{aligned} f(x_1, \ldots, x_n) = {} & x_i x_j \varphi_{ij}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \\ & + \psi_{ij}(x_1, \ldots, x_n) \end{aligned}$$

where the monomials of ψ_{ij} do not contain the product $x_i x_j$. Function φ_{ij} will be called the *second derivative* of f with respect to the variables x_i and x_j.

It can be shown (see [8] for instance) that f is supermodular if and only if, for any pair of variables x_i, x_j $(i \neq j)$ the second derivative φ_{ij} is *nonnegative* whatever the values assigned to the other variables.

In the quadratic case, this necessary and sufficient condition shows that f is supermodular if and only if all the coefficients of the quadratic terms are nonnegative. Observe that, if the coefficient of a linear term is nonnegative, then the corresponding variable may be definitely fixed to 1 in the search for a maximum[1]. Thus it is not restrictive to consider only quadratic supermodular functions in which all coefficients of the linear terms are strictly negative.

[1] Another interesting consequence of this property is that the maximum of any supermodular pseudoboolean function *without linear terms* is obtained by fixing all variables to 1.

Quadratic supermodular pseudoboolean functions thus belong to the class of pseudoboolean functions with linear terms of any sign and higher degree terms with nonnegative coefficients. These functions, termed *negative–positive* by some authors (after fixing to 1 the variables corresponding to positive linear terms) are obviously supermodular in view of the necessary and sufficient condition for supermodularity recalled earlier. It has been shown [18] that maximizing such functions can be performed polynomially via maximum network flow computations and the Ford–Fulkerson algorithm.

We provide here a new derivation of this (old) result by reducing the pseudoboolean maximization problem to the search for a maximum weight stable set on a bipartite graph (conflict graph).

Suppose we are interested in maximizing a *negative–positive* pseudoboolean function of the form

$$f(x_1,\dots,x_n)=\sum_{j=1}^{n} c_j\, x_j+\sum_{i=1}^{m} b_i \prod_{j\in N_i} x_j$$

where $c_j>0$ ($\forall j=1,\dots,n$) and $b_i>0$ ($\forall i=1,\dots,m$). Replacing x_j by $\bar{x}_j=1-x_j$ in all the linear terms, to get rid of the negative coefficients, we get:

$$f=-\sum_{j=1}^{n} c_j+\sum_{j=1}^{n} c_j\, \bar{x}_j+\sum_{i=1}^{m} b_i \prod_{j\in N_i} x_j \tag{1}$$

which is recognized as a *posiform* see [16].

Now, it is well-known that the maximization of such a posiform can be reformulated as the search for a maximum weight stable set in an associated graph $G(f)$ called the *conflict graph* (see Hammer [14]). Here the conflict graph $G(f)$ associated with (1) has node set $X(f)=V\cup W$ where $V=\{1,2,\dots,n\}$ is the set of nodes corresponding to linear terms, and $W=\{1,2,\dots,m\}$ the set of nodes corresponding to nonlinear terms, and there exists an edge linking two nodes if and only if one of the associated terms contains a variable, and the other the same variable complemented. Thus, $G(f)$ is a *bipartite graph* since it contains only edges joining nodes in V to nodes in W. The maximization of f thus reduces to determining a maximum weight stable set in the bipartite graph $G(f)$, where the nodes of V are assigned weights $c_j>0$ ($j=1,\dots,n$) and nodes of W are assigned weights b_i ($i=1,\dots,m$).

It is easily realized that the latter problem can be solved polynomially by computing a maximum flow between s and t in the transportation network deduced from $G(f)$ by:

– adding a node s (called source) and n arcs (s,j) with upper capacity bound c_j ($j=1,\dots,n$);

– adding a note t (called sink) and m arcs (i,t) with upper capacity bound b_i ($i=1,\dots,m$);

– assigning to the arcs (j,i) of $G(f)$ an upper capacity bound $+\infty$.

Let $[A,X(f)-A]$ where $s\in A$, $t\in\bar{A}=X(f)-A$ be the minimum cut obtained. The maximum weight stable set which we are looking for is then defined by:

$$S=(A\cap V)\cup(\bar{A}\cap W)$$

from which the values of the individual variables at the maximum of f are easily deduced.

Example. Consider the function

$$f=-2x_1-3x_2-2x_3-3x_4+2x_1x_2+3x_1x_3+6x_3x_4.$$

The associated transportation network and the corresponding maximum flow are indicated on Fig. 1 below. The nodes associated with the terms $3\bar{x}_2$, $3x_1x_3$ and $6x_3x_4$ form a maximum weight stable set in the bipartite graph. Its weight being 12, the maximum value of f is $12-(2+3+2+3)=2$ and is obtained for the set of values of the variables satisfying:

$$\bar{x}_2=1,\qquad x_1x_3=1,\qquad x_3x_4=1,$$

i.e. for $x_1=x_3=x_4=1$ and $x_2=0$.

3. Maximizing a cubic supermodular pseudoboolean function

Using the concept of conflict graph, we now present a new reduction to the maximum flow problem for the whole class of cubic supermodular pseudoboolean functions. We begin by stating a preliminary result characterizing cubic supermodular pseudoboolean functions.

Lemma 1. *A cubic pseudoboolean function $f(x_1,\dots,x_n)$ is supermodular, if and only if it can be written*

$$f(x_1,\dots,x_n)=-\sum_{j\in J^-}c_jx_j+\sum_{j\in J^+}c_jx_j+\sum_{i,j\in IJ}c_{ij}x_ix_j$$

$$+\sum_{(i,j,k)\in IJK^+}c_{ijk}x_ix_jx_k-\sum_{(i,j,k)\in IJK^-}c_{ijk}x_ix_jx_k+K \tag{2}$$

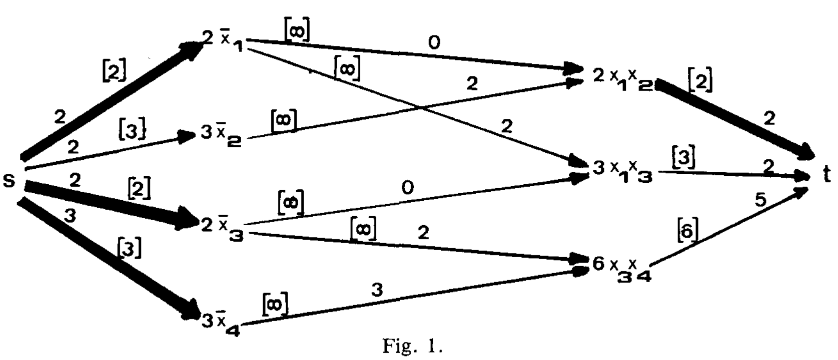

Fig. 1.

where the following conditions hold:

$$c_j \geq 0 \quad (\forall j \in J^-); \qquad c_j \geq 0 \quad (\forall j \in J^+); \qquad c_{ij} \geq 0 \quad (\forall (i,j) \in IJ);$$

$$c_{ijk} \geq 0 \quad (\forall (i,j,k) \in IJK^+); \qquad c_{ijk} \geq 0 \quad (\forall (i,j,k) \in IJK^-)$$

and

$$\forall (i,j) \in \{1,2,\dots,n\}^2,\ i<j: \quad c_{ij} \geq \sum_{k/(i,j,k) \in IJK^-} c_{ijk}$$

Proof. (i) *The condition is sufficient.*

For any $(i,j) \in \{1,2,\dots,n\}^2$, $i<j$, the second derivative is

$$\varphi_{ij} = c_{ij} + \sum_{k/(i,j,k) \in IJK^+} c_{ijk} x_k - \sum_{k/(i,j,k) \in IJK^-} x_k$$

which is nonnegative whatever the values of the variables if the conditions of the lemma hold.

(ii) *The condition is necessary.*

A cubic pseudoboolean function may always be put in the form:

$$f(x_1,\dots,x_n) = \sum_{j \in J^+} c_j x_j - \sum_{j \in J^-} c_j x_j + \sum_{(i,j) \in IJ} c_{ij} x_i x_j$$
$$+ \sum_{(i,j,k) \in IJK^+} c_{ijk} x_i x_j x_k - \sum_{(i,j,k) \in IJK^-} c_{ijk} x_i x_j x_k$$

with

$$c_j \geq 0 \quad (\forall j \in J^-); \qquad c_j \geq 0 \quad (\forall j \in J^+);$$

$$c_{ijk} \geq 0 \quad (\forall (i,j,k) \in IJK^+); \qquad c_{ijk} \geq 0 \quad (\forall (i,j,k) \in IJK^-).$$

For all $(i,j) \in \{1,2,\dots,n\}^2$ the second derivative is:

$$\varphi_{ij} = c_{ij} + \sum_{k/(i,j,k) \in IJK^+} c_{ijk} x_k - \sum_{k/(i,j,k) \in IJK^-} c_{ijk} x_k.$$

In order that φ_{ij} be nonnegative whatever the values taken by the variables, it is necessary that all $c_{ij} \geq 0$ (otherwise setting $x_k = 0$, $\forall k = 1,\dots,n$, φ_{ij} would be negative).

Now suppose that there exists $(i,j) \in \{1,2,\dots,n\}^2$ such that

$$c_{ij} < \sum_{k/(i,j,k) \in IJK^-} c_{ijk}$$

By setting $x_k = 0$ for all k such that $(i,j,k) \in IJK^+$ and $x_k = 1$ for all k such that $(i,j,k) \in IJK^-$, φ_{ij} would take a negative value. Thus all the conditions stated in the lemma are necessary. $\square$

In the following, we shall consider only cubic supermodular functions including all the possible linear terms with a negative coefficient (since in case of nonnegative linear terms, the corresponding variables can be fixed to 1, see remark in Section 2 above).

Now let us transform the cubic terms with negative coefficients in f by complementing all the variables. For a monomial $-x_i x_j x_k$ this provides:

$$-x_i x_j x_k = -1 - x_i x_j - x_i x_k - x_j x_k + x_i + x_j + x_k + \bar{x}_i \bar{x}_j \bar{x}_k .$$

Hence f may be rewritten as

$$f(x_1, \ldots, x_n) = \sum_{j \in J^+} c_j' x_j + \sum_{j \in J^-} c_j' x_j + \sum_{(i,j) \in IJ} c_{ij}' x_i x_j$$

$$+ \sum_{(i,j,k) \in IJK^+} c_{ijk} x_i x_j x_k + \sum_{(i,j,k) \in IJK^-} c_{ijk} \overline{x_i} \overline{x_j} \overline{x_k} + K' . \tag{3}$$

The nonnegativity of all coefficients of the nonlinear terms is obvious, except for c_{ij}'. Computing c_{ij}' we get:

$$c_{ij}' = c_{ij} - \sum_{k/(i,j,k) \in IJK^-} c_{ijk}$$

which is indeed nonnegative by virtue of Lemma 1.

We are now in a position to state:

Theorem 1. *The problem of maximizing a cubic supermodular pseudoboolean function can be reduced to a maximum weight stable set problem in a bipartite graph, hence can be solved polynomially by a maximum flow algorithm.*

Proof. We first observe that the number of terms in the transformed expression (3) of f is not greater than the number of terms in the initial expression (2). On the other hand, the conflict graph resulting from (3) is *bipartite* since, in each term, the variables appear either all in their direct form, either all in their complemented form. Thus, maximizing f is equivalent to a maximum weight stable set problem on a bipartite graph. □

Example. Consider the following cubic supermodular pseudoboolean function

$$\begin{aligned} f = & -6x_1 - 3x_2 - x_3 - 4x_4 \\ & + 5x_1 x_2 + 3x_1 x_3 + 2x_2 x_3 + 3x_1 x_4 + 4x_2 x_4 + x_2 x_3 x_4 \\ & - 2x_1 x_2 x_3 - 3x_1 x_2 x_4 . \end{aligned}$$

Let us convert this function into an homogeneous posiform.

We obtain:

$$\begin{aligned} f = & -6x_1 - 3x_2 - x_3 - 4x_4 \\ & + 5x_1 x_2 + 3x_1 x_3 + 2x_2 x_3 + 3x_1 x_4 + 4x_2 x_4 \end{aligned}$$

$$
\begin{aligned}
&+x_2x_3x_4\\
&+2(-1-x_1x_2-x_1x_3-x_2x_3+x_1+x_2+x_3+\bar{x}_1\bar{x}_2\bar{x}_3)\\
&+3(-1-x_1x_2-x_1x_4-x_2x_4+x_1+x_2+x_4+\bar{x}_1\bar{x}_2\bar{x}_4)\\
=&-x_1+2x_2+x_3-x_4\\
&+x_1x_3+x_2x_4\\
&+x_2x_3x_4\\
&+2\bar{x}_1\bar{x}_2\bar{x}_3+3\bar{x}_1\bar{x}_2\bar{x}_4\\
&-5\\
=&\bar{x}_1+2x_2+x_3+\bar{x}_4\\
&+x_1x_3+x_2x_4\\
&+x_2x_3x_4\\
&+2\bar{x}_1\bar{x}_2\bar{x}_3+3\bar{x}_1\bar{x}_2\bar{x}_4\\
&-7\\
=&f'-7.
\end{aligned}
$$

The maximum of f' is equal to the weight of the maximum weight stable set in the following conflict graph H (see Fig. 2) which is bipartite.

A maximum weight stable set of H can be found by computing a maximum flow between s and t in the transportation network of Fig. 3 (see Section 2).

A maximum flow is indicated on Fig. 3 above and gives a maximum weight stable set of the bipartite graph H; it is composed of the nodes associated with the terms $\bar{x}_1$, $2\bar{x}_1\bar{x}_2\bar{x}_3$, $3\bar{x}_1\bar{x}_2\bar{x}_4$ and $\bar{x}_4$. Its weight being 7 the maximum of f is $7-7=0$ and is obtained from the set of values of the variables satisfying:

$$\bar{x}_1=1, \qquad \bar{x}_1\bar{x}_2\bar{x}_3=1, \qquad \bar{x}_4=1, \qquad \bar{x}_1\bar{x}_2\bar{x}_4=1,$$

i.e. $x_1=x_2=x_3=x_4=0$.

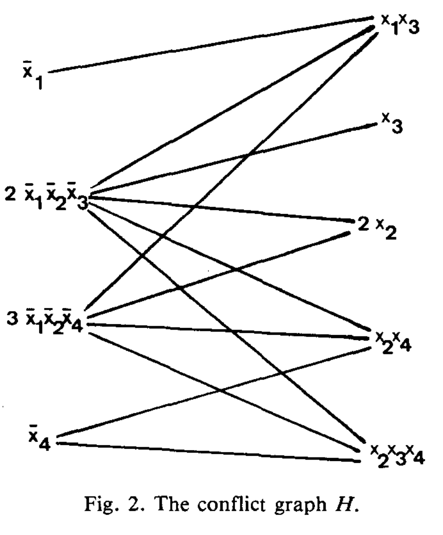

Fig. 2. The conflict graph H.

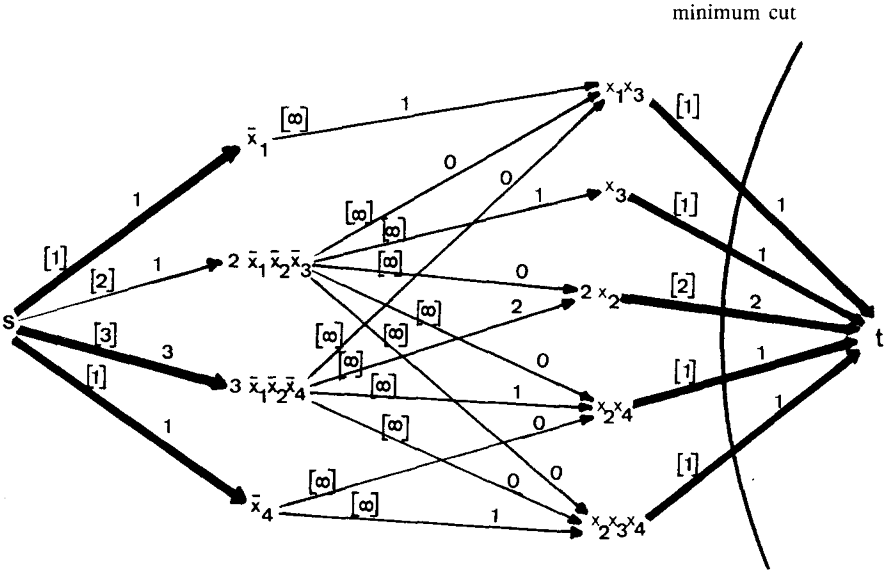

Fig. 3. The maximum flow on the transportation network associated with H.

4. Concluding remarks

We end up with a few remarks concerning the links between some important subclasses of pseudoboolean functions related to the results stated above:

(1) the subclass (S) of supermodular functions;

(2) the subclass (C) of pseudoboolean functions f which can be transformed into a negative–positive one by a change in variables (i.e. changing x_j to $1-y_j, y_j \in \{0,1\}$ in all terms of f for some subset of variables);

(3) the subclass (C') of pseudoboolean functions which can be converted by complementation of some variables into a homogeneous posiform, i.e. a posiform where each term contains either variables in their direct form only, or in their complemented form only (an homogeneous posiform has a conflict graph which is bipartite).

(i) *If a function f is in (C), then there is a function f' obtained from f by a change in variables such that f' is in (C').*

Indeed, since, once a function f in (C) has been converted into a negative–positive one, one clearly obtains an homogeneous posiform by complementing the variable in each linear term with negative coefficient (as indicated in Section 2).

(ii) *There exist functions in (C) which are not supermodular.*

To see this, we choose a negative–positive function where all quadratic terms are present, and we perform a change in variables $y_j = 1 - x_j$ for some (arbitrarily chosen) variables. In this transformation we obtain a function with negative quadratic terms, which obviously cannot be supermodular. Let us note that the quadratic functions in (C) have been considered by Hammer et al. [15] who call them *unimodular.*

(iii) $(C') \subset (S)$.

Let f be a function in (C') and consider a homogeneous posiform

$$f\colon \sum_U c(U) \prod_{i \in U} x_i + \sum_T c(T) \prod_{i \in T} \bar{x}_i$$

where $c(S)$ and $c(T) \geq 0$.

Each conjunction $\prod_{i \in U} x_i$ or $\prod_{i \in T} \bar{x}_i$ is supermodular (just consider the second derivatives). Since f is a sum of supermodular functions it is supermodular, too.

(iv) *There exist supermodular functions which are not in (C') and thus not in (C) either.*

The following function

$$\begin{aligned} f_1 = {} & 100x_1x_2 + 100x_2x_3 + 100x_3x_4 + 100x_1x_4 + 100x_1x_3 \\ & + 100x_2x_4 - x_1x_2x_3x_4 \end{aligned}$$

is clearly supermodular since all quadratic terms are present with a sufficiently large coefficient. However, the degree four term cannot be made positive by complementing all variables. Thus f_1 cannot be converted into a homogeneous posiform.

(v) *(S) and (C') coincide for cubic functions.*

This property follows immediately from (iii) and the results of Section 3.

From a practical and algorithmic point of view, the class (C') of pseudoboolean functions which can be converted into a homogeneous posiform is of much interest, since, the conflict graph of such functions being bipartite, they can be maximized via network flow techniques. Though it includes the subclass of negative–positive functions (modulo a change in variables), a proper characterization of membership in (C') remains to be found, except for the cubic case for which it has been shown here that supermodularity is a necessary and sufficient condition for a function to be in (C').

Note on related work

Upon completing this work, P. Hansen turned to our attention that our result had been independently obtained by L.A. Wolsey who kindly sent us his (unpublished) handwritten notes.

Unlike the approach presented here, his way of deriving the result doesn't make use of the conflict graph concept, but instead is based on an extension of the Balinski–Rhys property, which reads:

The problem

$$\text{Maximize } \sum c_i x_i + \sum_{|T| \geq 2} b_T \prod_{j \in T} x_i + \sum_{|T| \geq 2} \beta_T \left[1 - \prod_{j \in T} (1 - x_j) \right]$$

can be solved as a maximum network flow problem if $b_T \geq 0$ *and* $\beta_T \leq 0$.

It is then sufficient to observe that the transformation

$$x_i x_j x_k = 1 + x_i x_j + x_i x_k + x_j x_k - x_j - x_i - x_k - (1 - x_i)(1 - x_j)(1 - x_k)$$

applied to all cubic terms with negative coefficients in a supermodular cubic function leads to exactly the form above.

Acknowledgements

We thank P. Hansen and the referees for their constructive remarks and helpful comments on this paper.

References

[1] K. Aoshima and M. Iri, Comments on F. Hadlock's paper: finding a maximum cut of a planar graph in polynomial time, SIAM J. Comput. 6, (1) (1977) 86–87.
[2] C. Berge, Graphes et Hypergraphes (Dunod, Paris, 1970).
[3] S. Biesel-Guitonneau, Algorithmes gloutons et fonctions surmodulaires: théorie et application à un problème de sécurité dans les réseaux de Télécommunications, Ann. Télécommunications 35 (7–8) (1980) 266–271.
[4] A. Billionnet and I. Caradot, Comparaison expérimentale d'algorithmes pour les problèmes de recouvrement et de maximisation d'une fonction pseudo-booléene, RAIRO (série verte) 17 (1) (1983) 15–20.
[5] G.B. Dantzig, Linear Programming and Extensions (Princeton Univ. Press, Princeton, 1963).
[6] J. Edmonds, Paths trees and flowers, Canad. J. Math. 17 (1965) 449–467.
[7] S. Even, The maximal flow algorithm of Dinic and Karzanov: an exposition, MIT, LCS, TM 80 (1976).
[8] M.L. Fisher, G.L. Nemhauser and L.A. Wolsey, An analysis of approximations for maximizing submodular set functions I, Math. Programming 14 (1978) 265–294.
[9] L.R. Ford and D.R. Fulkerson, Flows in Networks, (Princeton Univ. Press, Princeton, 1962).
[10] M.R. Garey and D.S. Johnson, Computers and Intractability: An Introduction to the Theory of NP-completeness (W.H. Freeman, San Fransisco, 1979).
[11] M. Gondran and M. Minoux, Graphes et Algorithmes, Collection de la Direction des Etudes et

Recherches d'Electricité de France (Eyrolles, Paris, 1979).

[12] Grötschel, L. Lovasz and A. Schrijver, The ellipsoïd method and its consequences in combinatorial optimization, Combinatorica I (2) (1981) 169–197.

[13] F. Hadlock, Finding a maximum cut of a planar graph in polynomial time, SIAM J. Comput. 4 (3) (1975) 221–225.

[14] P.L. Hammer, The conflict graph of a pseudoboolean function, Bell Labs. Operations Research Center (W. Long Branch, New Jersey, 1978).

[15] P. Hammer, P. Hansen and B. Simeone, Roof duality, complementation and persistency in quadratic 0–1 optimization, Math. Programming 28 (1984) 121–155.

[16] P. Hansen, Quelques approches de la programmation non linéaire en variables 0–1.

[17] M. Minoux, Accelerated greedy algorithms for maximizing submodular set functions, 8th IFIP TC 7 Optimization Conference (Springer, Berlin, 1977) 234–243.

[18] J.M.W. Rhys, A selection problem of shared fixed costs and network flows, Management Sci. 17 (3) (1970) 200–207.